Gravitational Lensing of Galaxies with a Hierarchical Dirichlet Process Prior

CASIS 18th Annual Signal & Image Sciences Workshop
May 21, 2014

Michael D. Schneider
Physics Division, LLNL

Lawrence Livermore National Laboratory

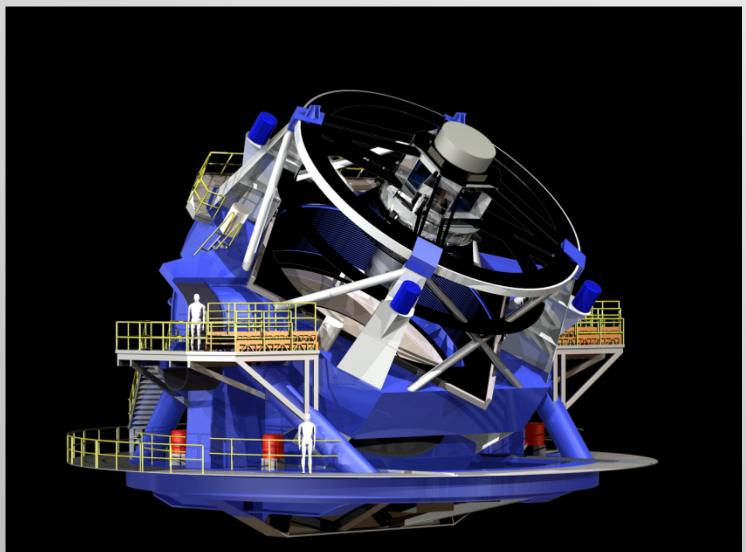
LLNL-PRES-654412
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

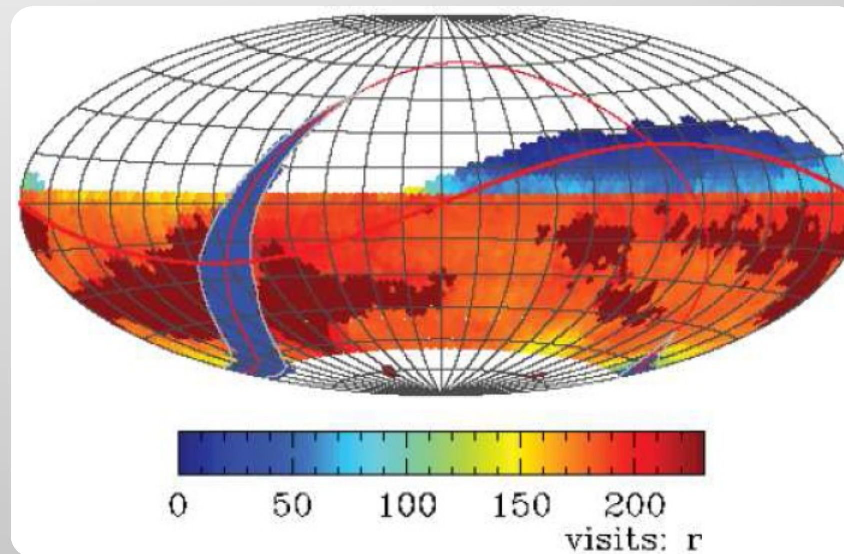
Large Synoptic Survey Telescope: A Deep, Wide, Fast, Uniform Sky Survey

Construction start: next month!   First light: 2020

8.4m telescope 18,000+ deg^2 10mas astrom. r<24.5 (<27.5@10yr)

ugrizy   0.5-1% photometry



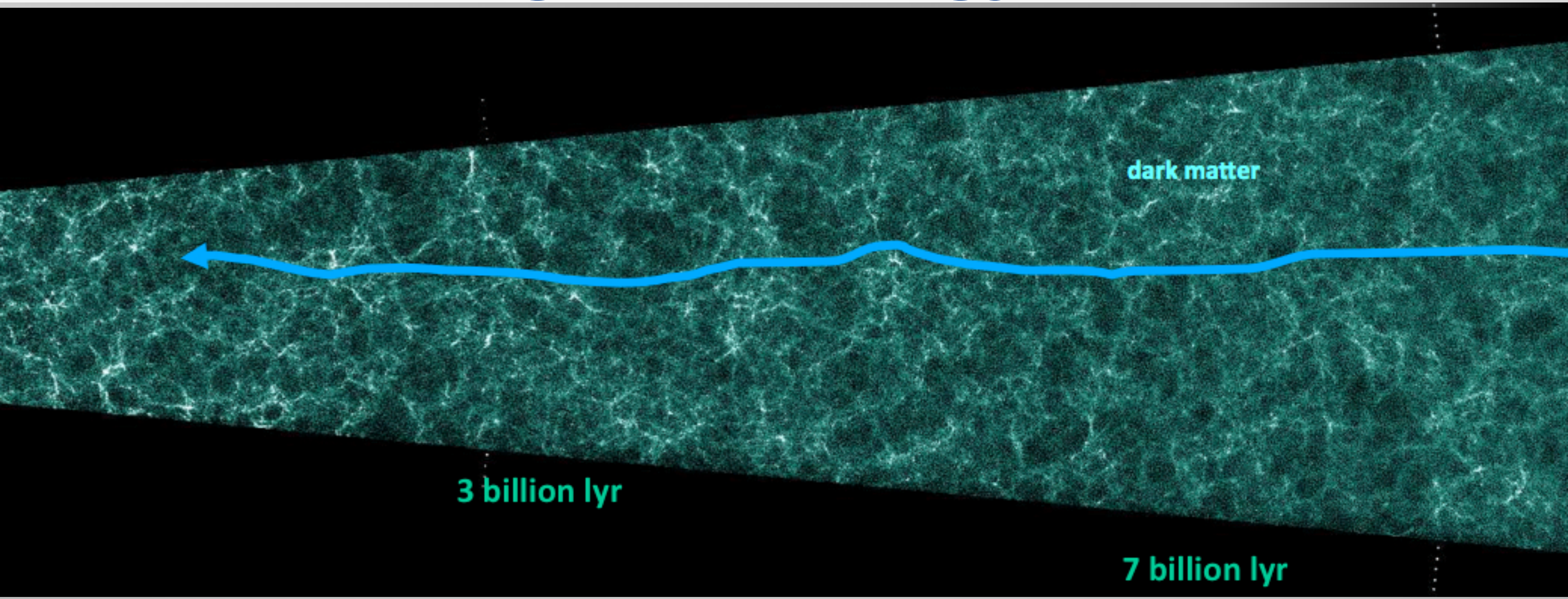


3.2Gpix camera 2x15sec exp/2sec read   **15TB/night**   **20 B objects**

Imaging the visible sky, once every 3 days, for 10 years (825 revisits)

Credit: T. Tyson

Gravitational lensing traces mass structure vs cosmic time – promising dark energy probe

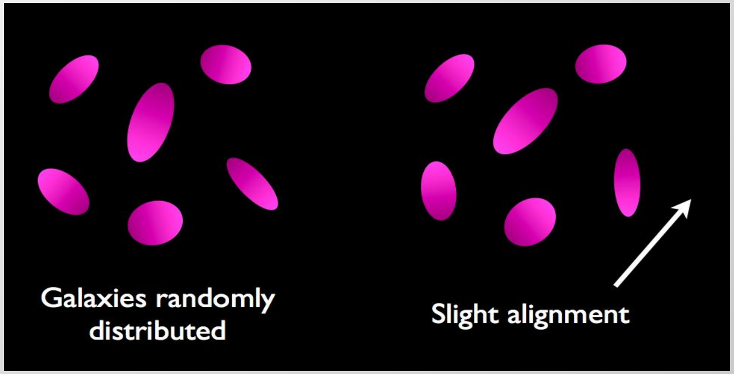


LSST will give positions, shapes, and distance estimates of 4 billion galaxies.

Credit: Bard, Meyers (SLAC)

But to measure gravitational lensing, we need to measure galaxy shapes...

- Lensing is a small distortion (<1%)
- Distortions due to atmosphere, telescope, CCD, measurement method are large (>~10%)!



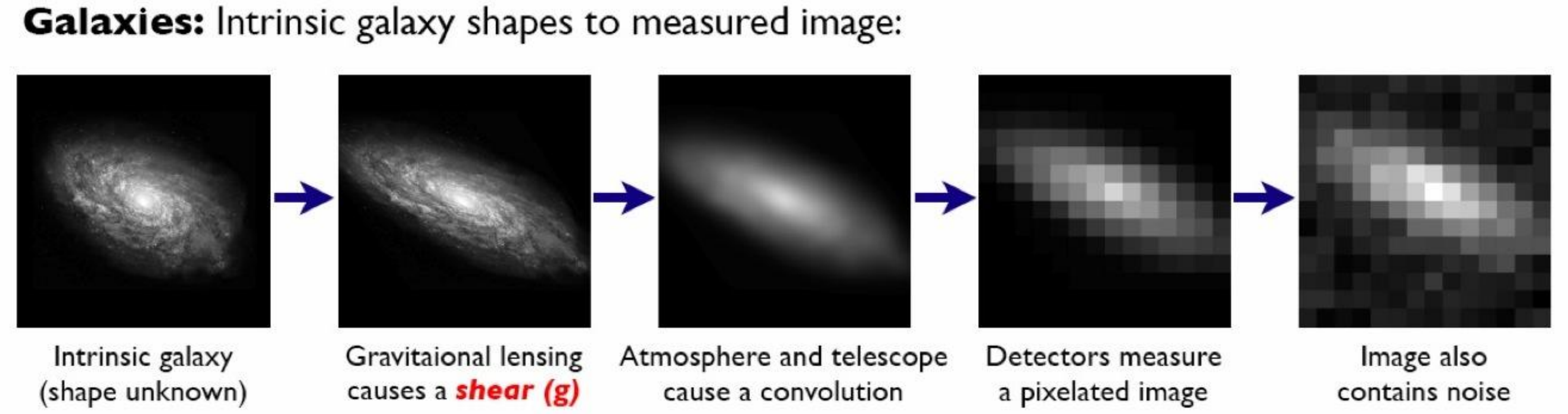
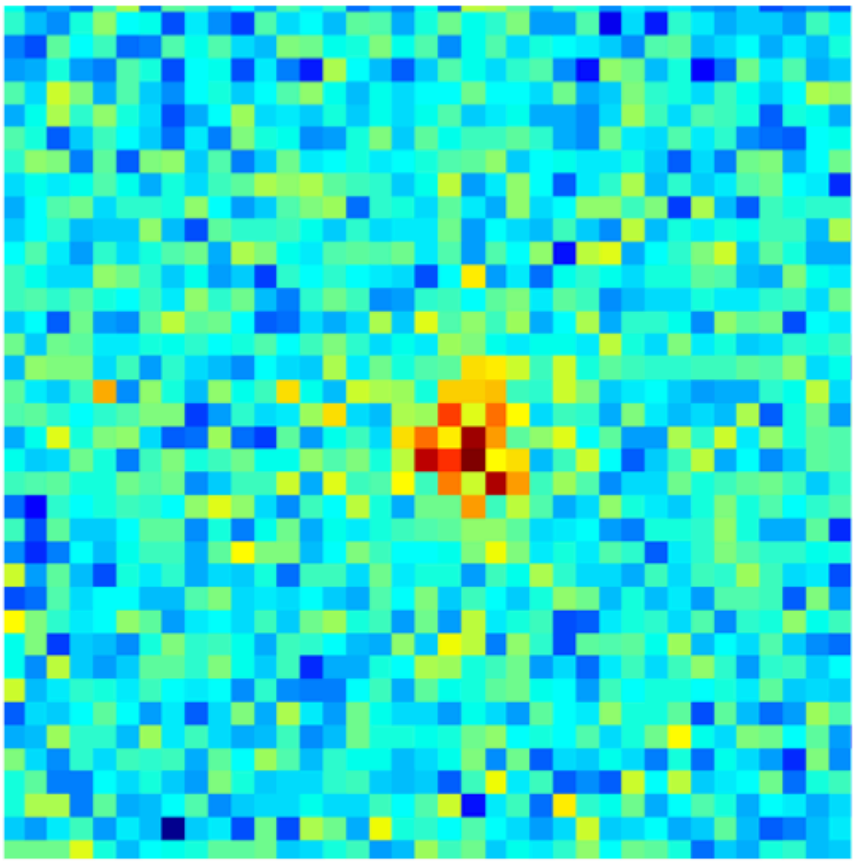


Image credit: S. Bridle

A typical galaxy image for cosmic shear



Intrinsic galaxy shape
b/a ~ 0.5

Uncertainty due to noise
σb/a ~ 0.5

Modification due to lensing
Δb/a ~ 0.05

Effect of changing w by 1%
δb/a ~ 0.0005

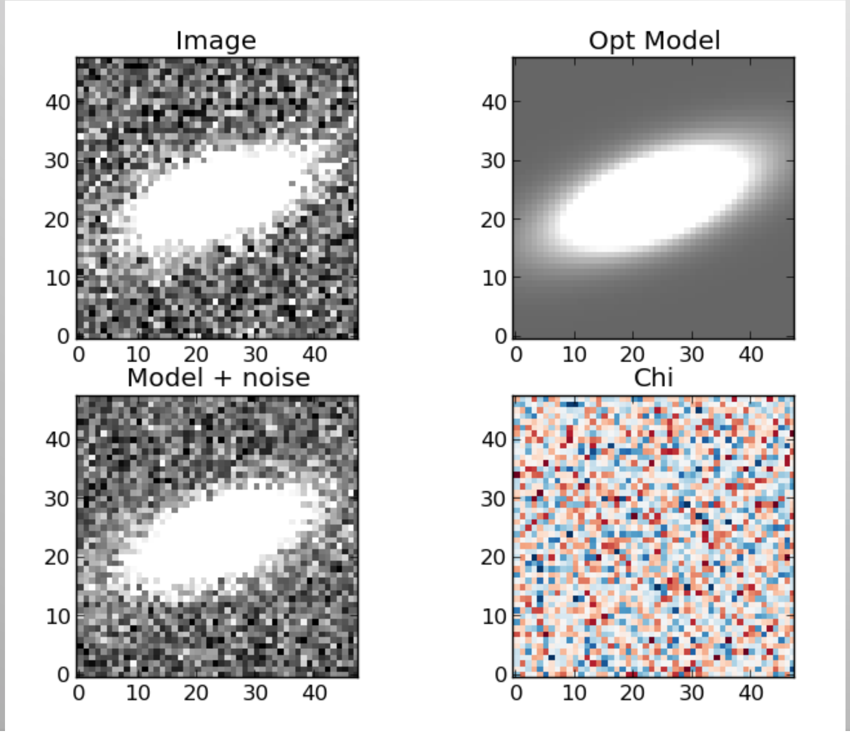
Credit: S. Bridle

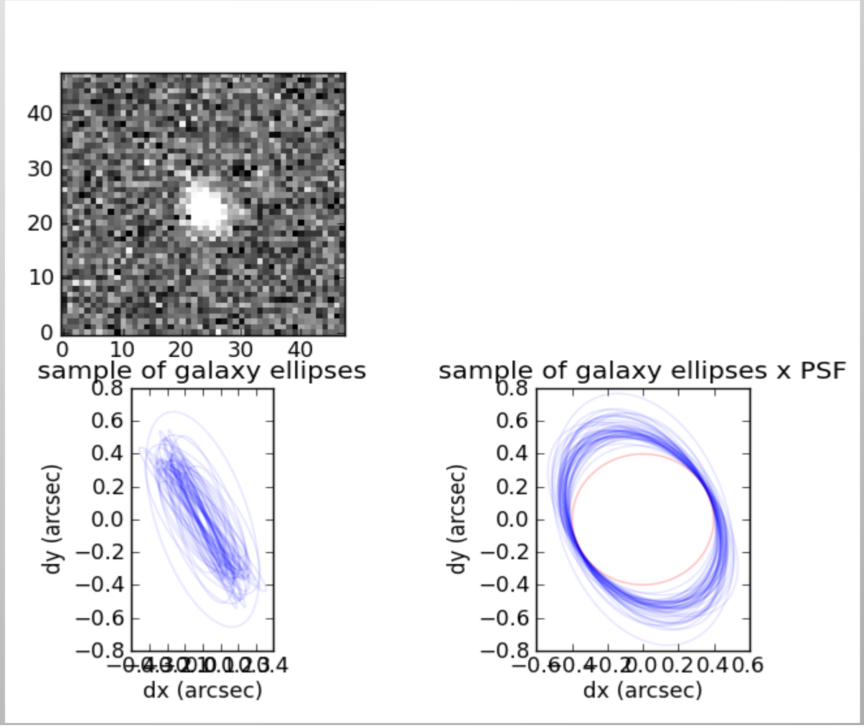
Credit: Bard, Meyers (SLAC)

The Tractor

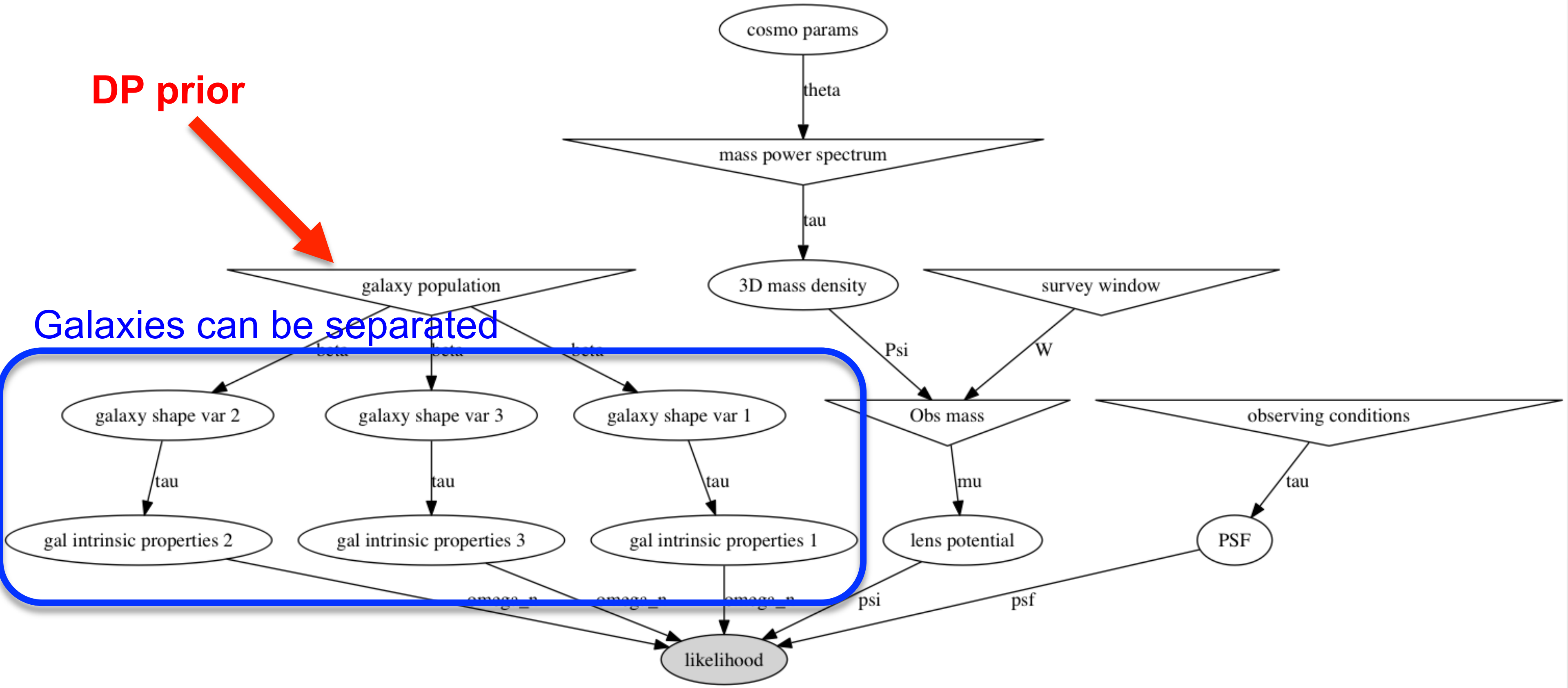
(D. Lang and D. Hogg)

- A generative model galaxy fitting framework framework
- Models PSF and galaxies as mixtures of Gaussians, to perform all convolutions analytically
- Light profiles modeled as constrained mixture of Gaussians
- Start with simple, elliptically-symmetric models
- Easy to take samples from the posterior probability distribution





Our statistical model



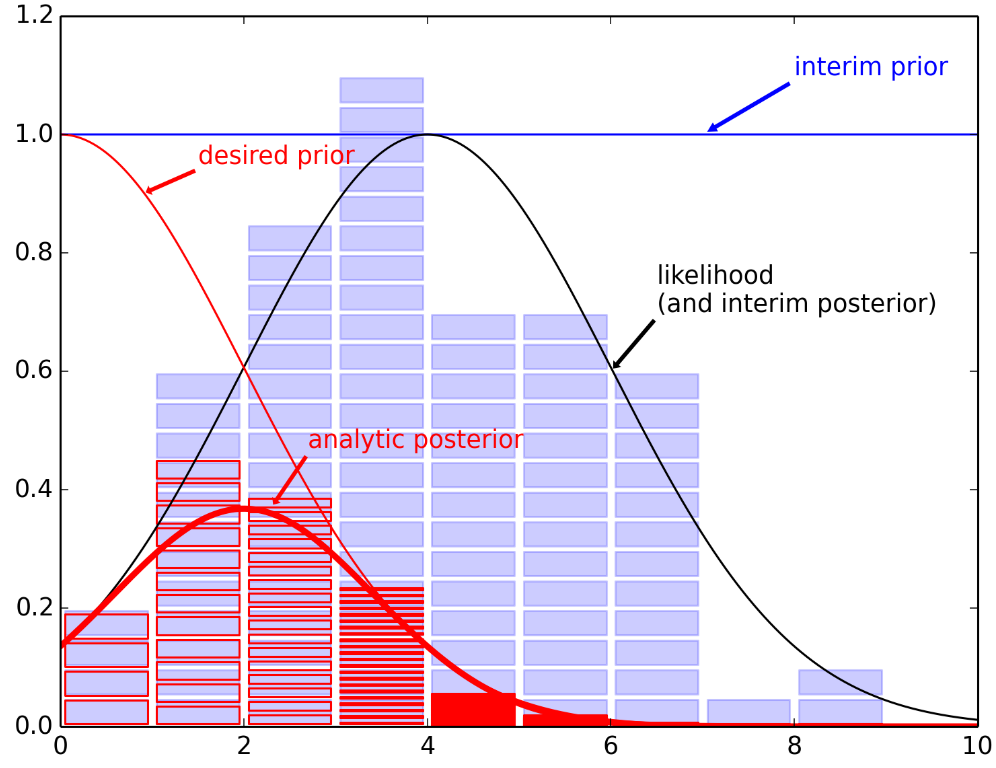
Hierarchical Inference by Importance Sampling

Use Tractor to sample galaxies individually, using interim prior.

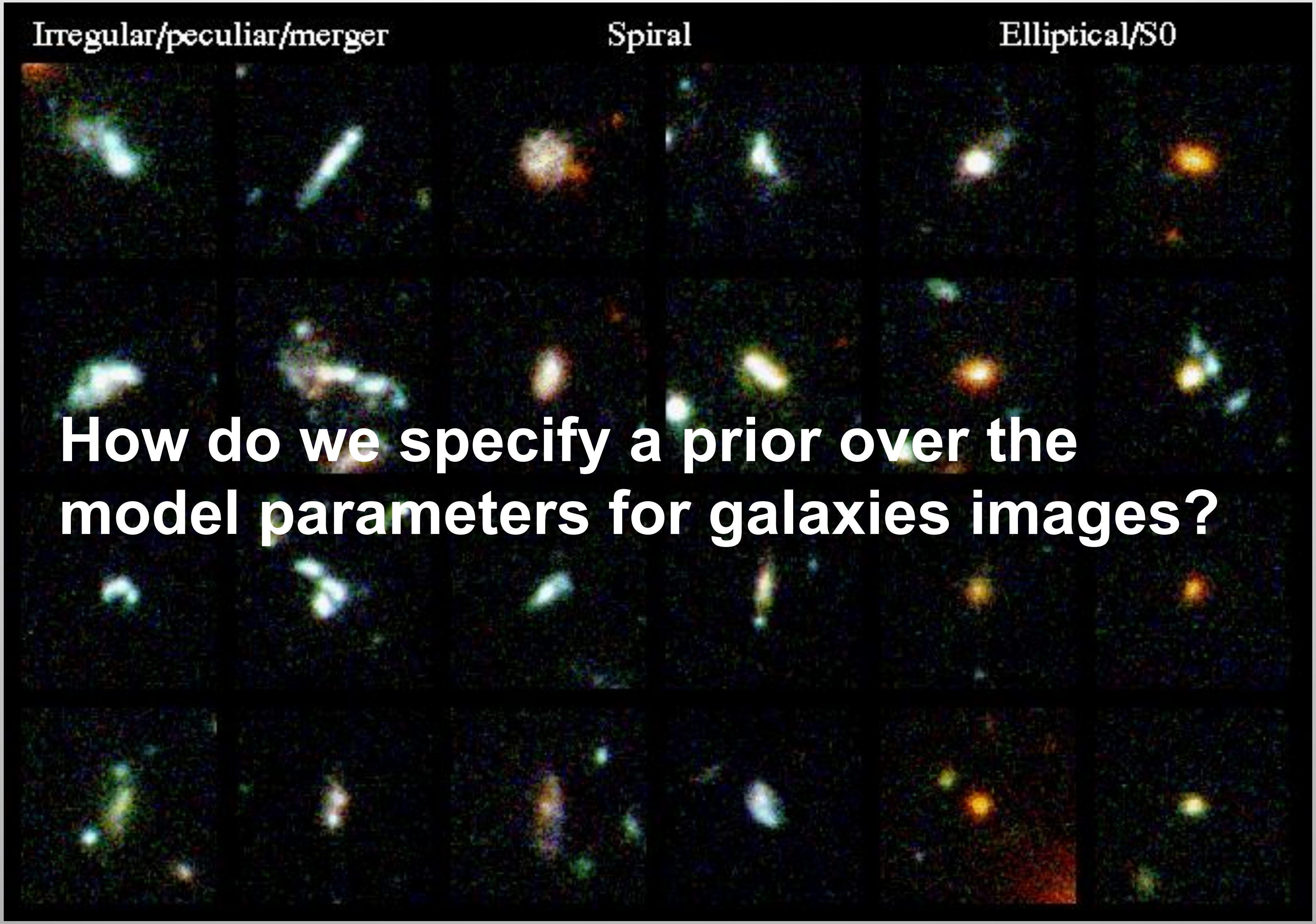
Importance sampling re-weights those samples to give desired posterior.

The inference of the independent galaxy shapes is *massively parallelizable*

Drawback is inefficiency and increased noise if interim prior is not carefully chosen.



Real galaxy images are diverse



How do we specify a prior over the model parameters for galaxies images?

Dirichlet Process Mixture Models

Example from Wang & Dunson (2011)

- Consider a mixture of Normal distributions,

$$y_i \sim N(\tilde{\mu}_i, \tilde{\tau}_i^{-1}), \quad (\tilde{\mu}_i, \tilde{\tau}_i^{-1}) \sim P, \quad i = 1, \ldots, n, \quad P \sim DP(\kappa P_0),$$

$\boldsymbol{\alpha}_i \equiv (\tilde{\mu}_i, \tilde{\tau}_i^{-1})$ parameters for galaxy i

P_0 DP *base* distribution

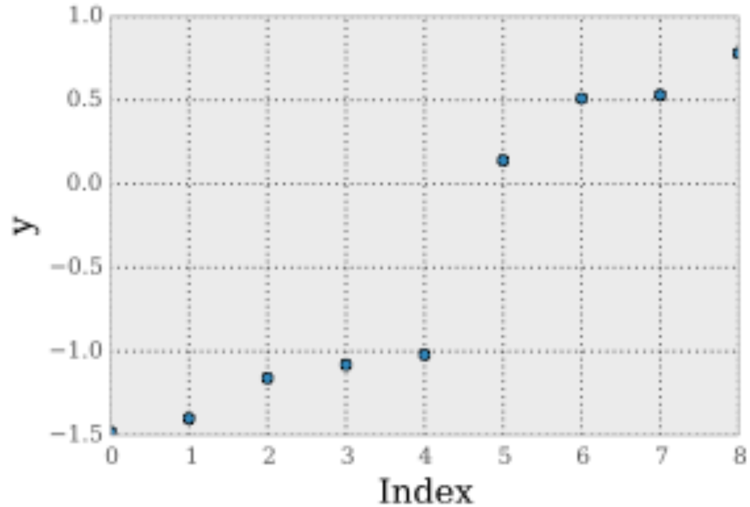
κ DP *precision* parameter

- By marginalizing over the probability measure *P* for each mixture we get the DP prediction rule (Blackwell & McQueen 1973),

$$(\boldsymbol{\alpha}_i | \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}) \sim \left(\frac{\kappa}{\kappa + i - 1}\right) P_0 + \left(\frac{1}{\kappa + i - 1}\right) \sum_{j=1}^{i-1} \delta_{\boldsymbol{\alpha}_j}.$$

EXAMPLE FROM NEAL (2000)

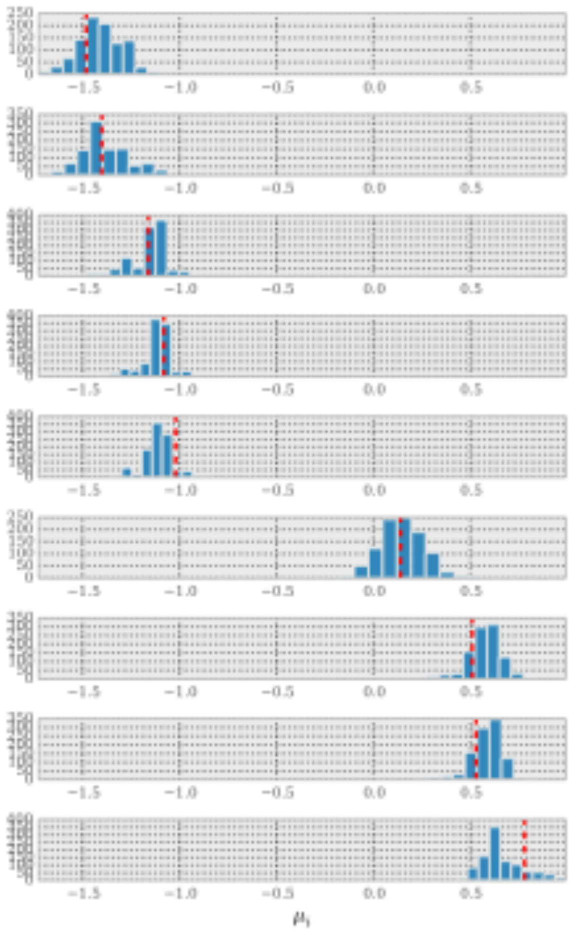
From Neal (2000), consider a 1-D data set, y, with 9 values.



Fix the precision $\tau = 100$ and let $\boldsymbol{\alpha}_i = \mu_i$ be the sampling parameter.
The DP prior is specified with a base distribution $P_0 = N(0, 1)$ and parameter $\kappa = 1$.

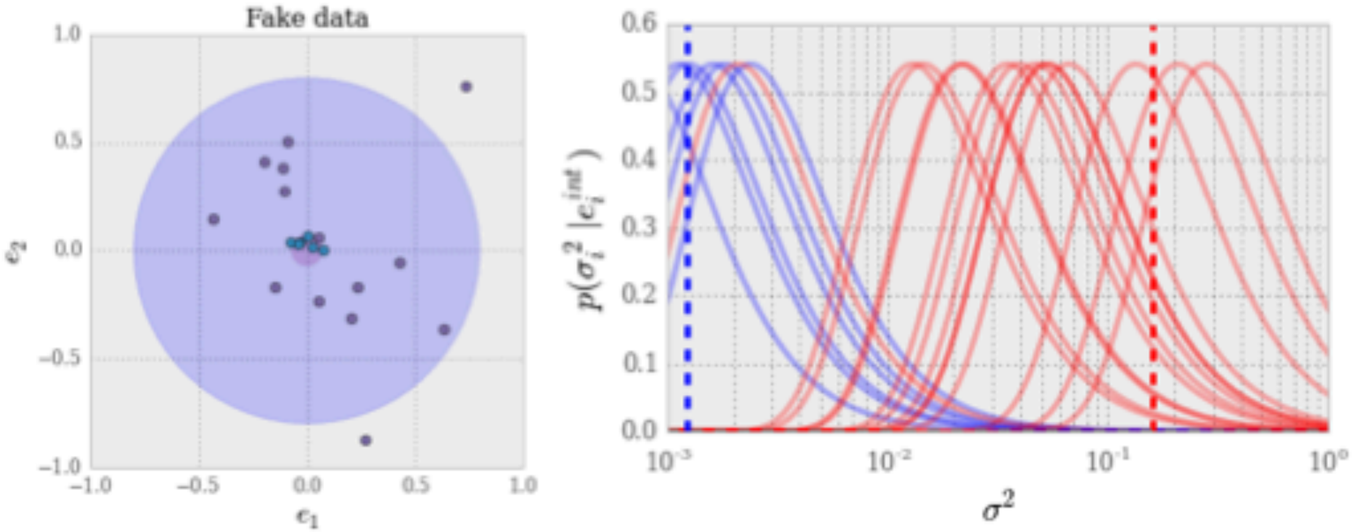
Example from Neal (2000) – Posterior samples show clusters

- Using conjugate Normal for the DP prior base distribution allows *Gibbs sampling*.
- Right: Histograms of posterior samples of μ_i for $i = 1, \ldots, 9$.
- Red lines: y_i
- 4 clusters are found in the data – note the histogram peaks line up among clusters.



DP PRIOR FOR $\sigma^2_{e^{\mathrm{int}}}$ (3)

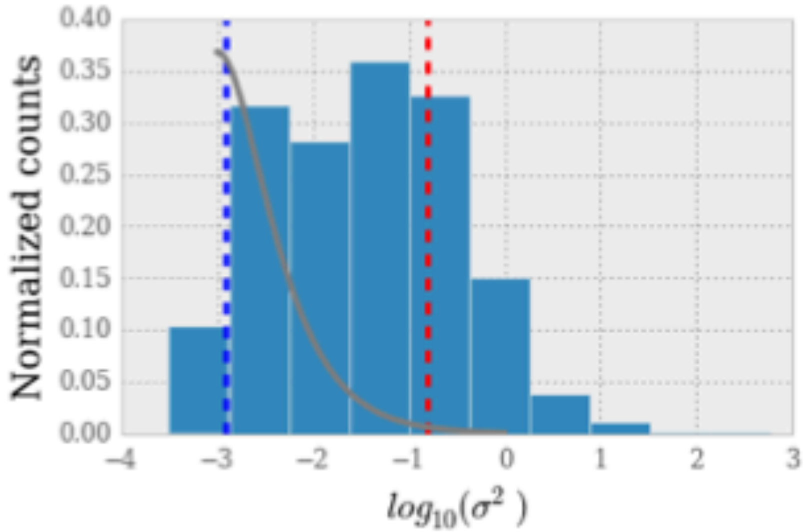
To show off the DP and to mimic the diversity in galaxy populations, generate fake data from a mixture of 2 Normal distributions with different widths (blue vs purple points in the left panel).



The right panel shows the posterior for σ_i^2 given only a single galaxy observation. Vertical lines indicate values used to generate the data. Red: galaxies from the 'wide' population, blue: galaxies from the 'narrow' population.

DP prior for $\sigma^2_{e^{\mathrm{int}}}$ (4)

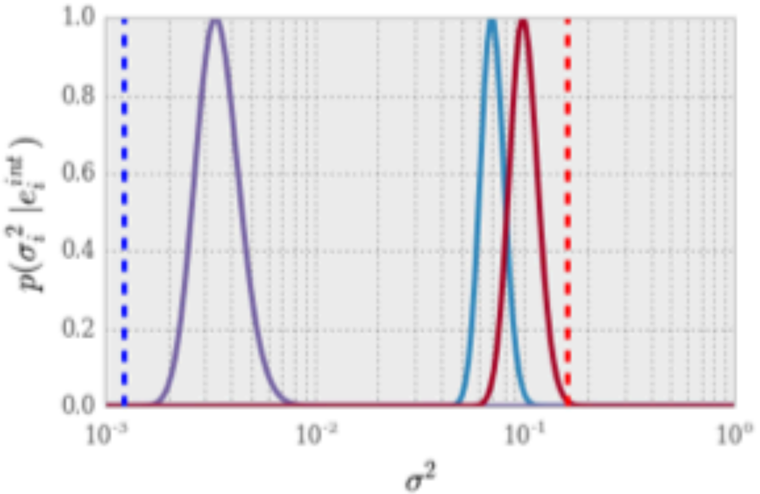
We can estimate a marginal posterior for σ^2 with a histogram of all samples for all galaxies. **Note: the 2 input values are correctly inferred.**



The red and blue vertical lines show the 2 values used to generate the data from the mixture model. The grey line shows P_0 as a conjugate inverse-Gamma distribution (not a very efficient choice).

Wrong inference: assume single Gaussian

- We could easily draw catastrophically wrong conclusions by assuming a less flexible prior.
- Assume the data generated from a single Normal with mean 0 and variance σ^2.
- What posterior would we infer for σ^2?
  - Will be too narrow $\rightarrow$ overconfidence in our result.
  - Sub-optimal: some data assumed to have larger variance than truth.



- Blue: all galaxies
- Red: 'broad' pop. only
- Purple: 'narrow' pop. only
- vertical lines: true σ^2 values
- All posteriors too narrow - exclude truth in all cases.

Non-conjugate priors

- Most DP sampling methods in the literature are limited by requirements on conjugacy.
- By importance sampling, we can use non-conjugate priors in the DP model with little additional computational complexity.

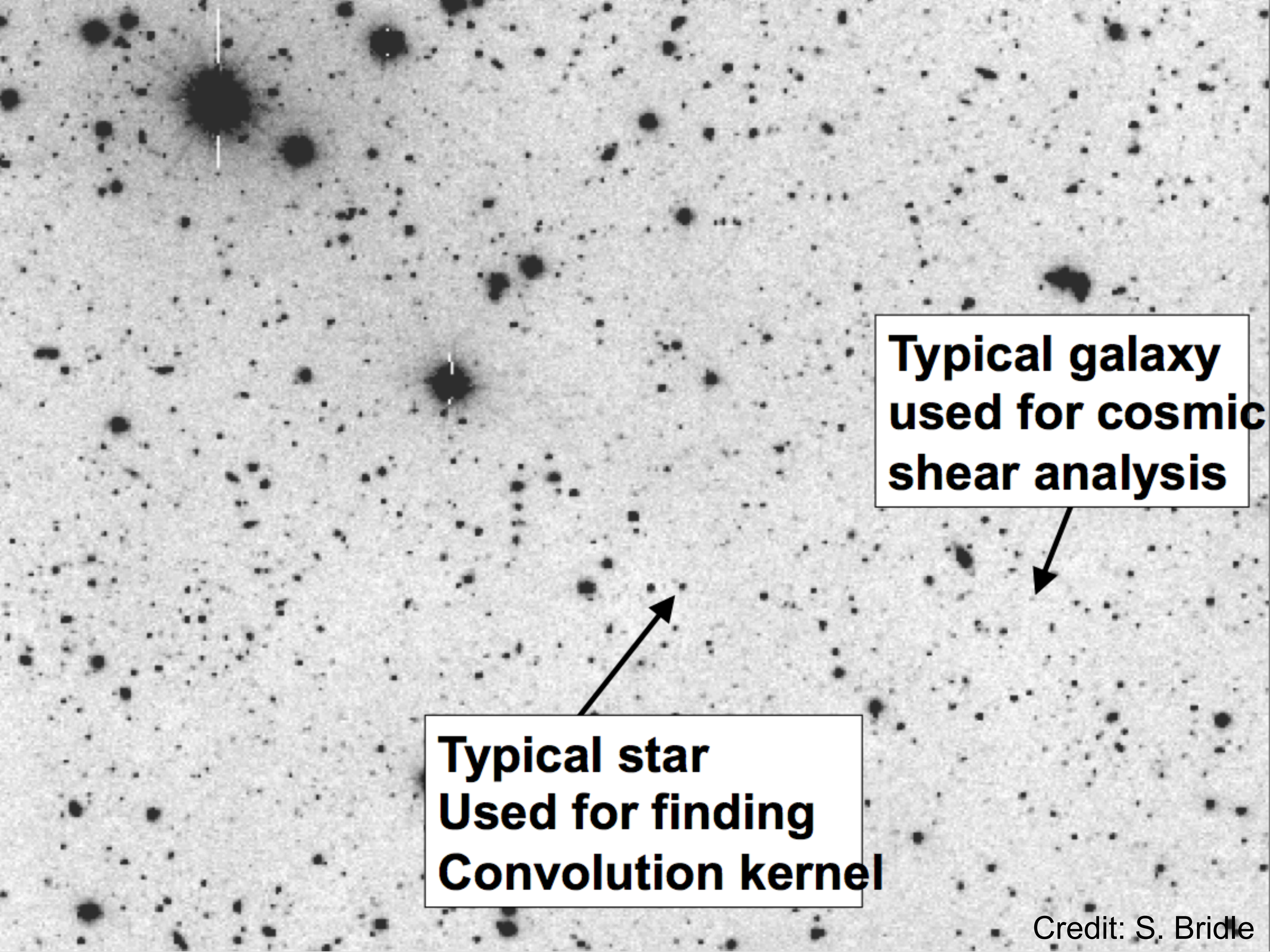
Summary: Benefits of DP prior

- Avoid overconfidence due to wrong prior specification.
- Find natural groupings of galaxy properties.
  - *Learn* how galaxies are formed – ancillary science.
- Sample based on constraints from other subsets of the data -> more efficient.

Challenges:

- The sampling model includes parameters for every galaxy.
  - Need to scale for billions of observed galaxies.
  - Parallelized Gibbs sampling?

Lawrence Livermore
National Laboratory



Typical galaxy used for cosmic shear analysis
Typical star Used for finding Convolution kernel
Credit: S. Bridle

Our statistical model

- Joint PDF factors as:

$$P(D, e^{\rm obs}, e^{\rm int}, \sigma_e, g)$$
$$= P(D|e^{\rm obs})P(e^{\rm obs}|e^{\rm int}, g)P(e^{\rm int}|\sigma_e)$$
$$P(g)P(\sigma_e)$$